Characterization of Replicative and Error-Prone DNA Polymerases

A Senior Honor's Thesis

Presented in Partial Fulfillment of the Requirements for Graduation with Research Distinction in Biochemistry in the Undergraduate Colleges of the Ohio State University

Walter Zahurancik

The Ohio State University

June 2012

Project Advisor: Dr. Zucai Suo, Department of Biochemistry

Contents

Introduction

The faithful replication of DNA is a process that is essential to the survival of an organism. Many enzymes are involved in the replication of DNA, some of which are responsible for the general replication of the genome and others which are responsible for repairing damaged DNA. DNA is frequently prone to damage caused by ionizing radiation, ultraviolet light, chemical modification, and spontaneous forms of damage such as depurination [1]. Understanding how these enzymes function is the key to understanding how DNA is consistently replicated faithfully.

Three DNA polymerases I have been working with are DNA Polymerase II (Dpo2) and DNA Polymerase III (Dpo3) from *Sulfolobus solfataricus*, and human DNA Polymerase ε. Dpo2 and Dpo3 have been characterized as error-prone DNA polymerases [2]. That is, they are capable of bypassing various types of damaged DNA when replicative polymerases fail, often due to having larger active sites which are capable of interacting with the unusual geometry of modified nucleotides [3]. Additionally, the mechanisms of action of such enzymes may be error-inducing in themselves. For instance, a repair process may result in a mutagenic frameshift caused by nucleotide insertion or deletion, which is catastrophic to DNA stability [4]. Characterization of these two enzymes will help us to better understand the mechanisms behind lesion bypass as well as the degree to which they preserve the integrity of the genetic code.

Human DNA polymerase ε is the primary enzyme responsible for leading strand synthesis in the human genome replication system. As a replicative polymerase, Pol ε is fast and highly processive, which are features necessary for a protein charged with replicating such a vast genome [5].

In my thesis, I will discuss several ongoing projects I am currently undertaking in the lab. First, I will examine circular dichroism spectroscopy experiments I have performed on Dpo2 and Dpo3. I will discuss the results of thermostability studies and compare them to previously published results and examine any discrepancies. Second, I will discuss our approaches to the purification of an exonuclease-deficient mutant of human DNA polymerase ε. I will begin by reviewing an experiment that was performed to optimize protein expression. This will be followed by an examination of the purification procedure used to obtain the exonuclease-deficient Pol ε. Furthermore, I will discuss kinetic assays that were performed using wild-type Pol ε. In the final section, I will explain the rationale behind the design of primers that will be used to obtain an exonuclease-deficient mutant of Dpo2 and an improved exonuclease-deficient Pol ε mutant. In particular, I will highlight the difficulty in determining which residues should be modified in Dpo2 and why the currently published "exonuclease-deficient" Pol ε retains some exonuclease activity.

I. Circular Dichroism Analysis of DNA Polymerase II and III from *Sulfolobus solfataricus*

Background

Circular dichroism spectroscopy is a technique that takes advantage of the chiral properties of macromolecules to produce a unique spectrum from which the general secondary structure and other unique properties can be derived. CD spectroscopy utilizes circular polarized light in the analysis of macromolecules. Circularly polarized light has both left-handed and right-handed components. Thus, circularly polarized light itself is chiral. As such, circularly polarized light reacts differently with different chiral molecules. Depending on the chirality of the molecule, the molecule will absorb more of either the left-handed component or the right-handed component of the circularly polarized light. The difference between the absorbance of the two components yields the circular dichroism spectrum.

The circular dichroism spectrum is plotted as signal (in units of molar ellipticity) versus wavelength. The different types of secondary structure typically generate a characteristic spectrum. For instance, alpha helices tend to have high positive absorbance values from 190-200 nm and a strong negative absorbance from 200-235 nm. Beta sheets absorb at wavelengths similar to those absorbed by alpha helices, although the magnitudes of absorbance are not as high. Finally, random coils show a strong negative absorbance from 190-210 nm (Figure 1). From the molar ellipticity values obtained at each wavelength for a given protein, the secondary structure content of a protein can be estimated.

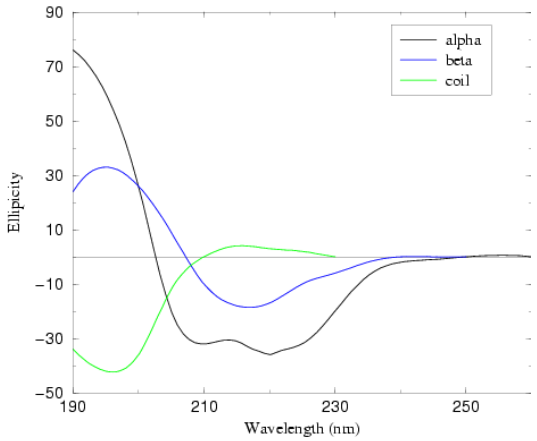

Figure 1. Characteristic spectrum of the three types of protein secondary structure.

In addition to secondary structure, protein stability may be analyzed by varying temperature and taking measurements at a constant wavelength. As temperature increases, kinetic energy increases. Consequently, as the fluid motions of the protein become faster and faster, the various interactions that dictate the protein's secondary and tertiary structures become increasingly weaker until they are broken altogether. On a CD spectrum, this transformation is often characterized by a sigmoidal curve. The melting transition of a protein is marked by a sharp increase in signal, followed by steady and static step. The melting temperature, or T_m, is the midpoint of the transition. For some proteins, the curve may have multiple steps which are an indication that there may be multiple stable intermediates that form during the protein folding and unfolding processes (Figure 2).

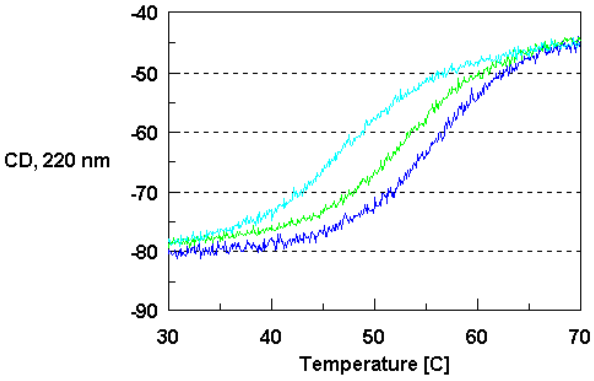

Figure 2. Generic CD spectrum showing temperature dependent studies.

Methods

Circular dichroism spectroscopy was utilized previously in our lab in the identification of unfolding intermediates for DNA Polymerase IV (Dpo4) from *S. solfataricus* [6]. As a thermophilic organism, *S. solfataricus* produces thermostabile enzymes that are capable of withstanding the high physiological temperature of *S. solfataricus* [7]. Our lab optimized conditions for circular dichroism spectroscopy analysis and had determined that Dpo4 has a three-state cooperative unfolding trend with melting temperatures at 89.3 and 102.6 [6]. Using the circular dichroism spectroscopy studies of Dpo4 as a model, we conducted brief CD spectroscopy analysis of DNA Polymerase II (Dpo2) and DNA Polymerase III (Dpo3) from *S. solfataricus*.

For circular dichroism spectroscopy studies of Dpo4, buffer conditions were optimized to 25 mM Na_3PO_4 pH 7.5, 50 mM NaCl, and 5 mM $MgCl_2$. Using a centrifugal concentrator, Dpo2 and Dpo3 were buffer exchanged into conditions identical to those used in the analysis of Dpo4. Using BCA assays, the concentrations of Dpo2 and Dpo3 were determined to be 6.5 μM and 17 μM, respectively.

During initial wavelength scans of Dpo3, the xenon lamp dynode voltage would become higher than 850 V when scanning wavelengths between 200 and 210 nm. This causes the scan to stop unless voltage drops below 850 V. High dynode voltage can be caused by high salt concentrations and high protein concentrations. This is problematic because valuable data regarding secondary structure can be obtained in this range of wavelengths. To optimize signal detection and prevent the dynode voltage from increasing beyond the maximum threshold of 850 V too quickly, various concentrations of Dpo3 were scanned at 16 °C. Dpo3 was serially diluted to concentrations of 4 μM, 2 μM, 1μM, 500 nM, 250 nM, and 100 nM. In this range of concentrations, wavelength scans from 300-190 nm finished between 190 and 193 nm which is negligible variation for our purposes as wavelengths under 200 nm are not particularly valuable in the determination of secondary structure content. The signal with the greatest magnitude was produced for the 4 μM sample, so 4 μM protein samples were used for the remainder of all Dpo2 and Dpo3 scans.

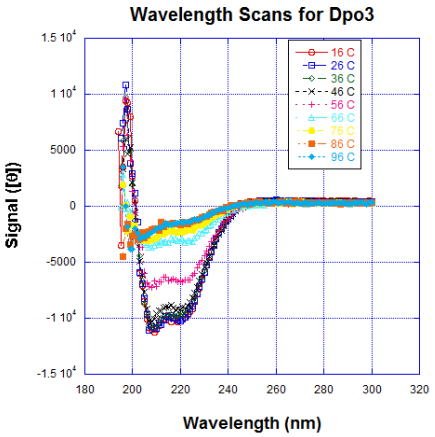

Figure 3. Wavelength dependent scans for Dpo3 at various temperatures.

Wavelength scans were performed over a range of 300-190 nm, with averaging times of 5 seconds and data points collected every 1 nm. Temperatures were held constant for each scan and multiple scans were performed at different temperatures to generate a series of curves that illustrate the gradual change in protein secondary structure

with temperature variance. Scans were performed between 16 °C and 106 °C at every 10 °C (Figure 3).

For temperature dependent scans, a constant wavelength of 208 nm was selected because 208 nm, on average, gave the signal with the greatest magnitude in the wavelength dependent scans. Temperature was varied from 16-106 °C and data was collected every 1 °C. Averaging time was held at 5 seconds and temperatures were allowed to equilibrate for 30 seconds before data was collected (Figure 4). An additional temperature scan was performed at 222 nm for Dpo2, but did not produce a noticeably different result from the scan at 208 nm (data not shown).

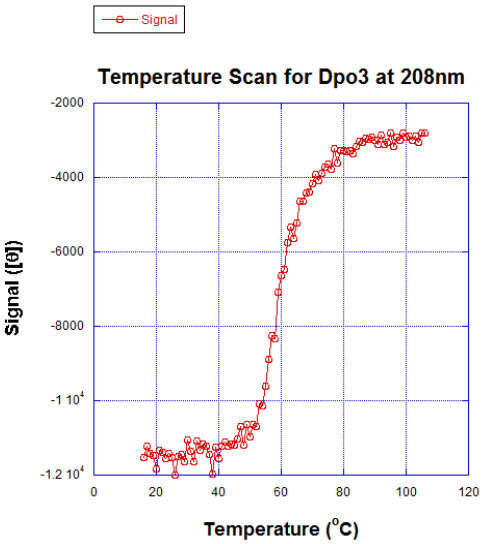

Figure 4. Temperature dependent scan for Dpo3 at a constant wavelength of 208 nm.

Wavelength dependent scans for Dpo2 were performed in a manner identical to those for Dpo3 (Figure 5). The temperature dependent scan for Dpo2 was performed at a constant wavelength of 208 nm with temperatures varying from 16-106 °C. Measurements were taken every 2 °C, with averaging times of 5 seconds and equilibration times of 30 seconds. The temperature scan for Dpo2 was also allowed to run in the reverse direction immediately following completion of the forward direction scan to detect the possibility of protein renaturation. Beginning at 106 °C, the temperature was gradually lowered back to 16 °C, with measurements taken every 2 °C (Figure 6).

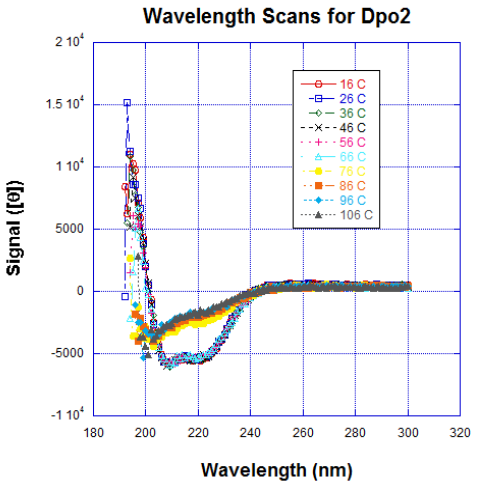

Figure 5. Wavelength dependent scans for Dpo2 at various temperatures.

Results

The circular dichroism spectra for the Dpo3 wavelength scans (Figure 3) show little change between 16 °C and 46 °C. At 56 °C, and again at 66 °C, there is significant increase in molar ellipticity. Molar ellipticity then continues to increase slightly at 76 °C and 86°C. There is no discernible difference between the spectra at 86 °C and at 96 °C. The temperature dependent spectra for Dpo3 (Figure 4) shows a relatively stable molar ellipticity between 16 °C and 46 °C. At roughly 50 °C, molar ellipticity begins to increase sharply. Between 65 °C and 70 °C, the increase in signal becomes less dramatic. Finally, around 80 °C, the signal once again becomes relatively stable.

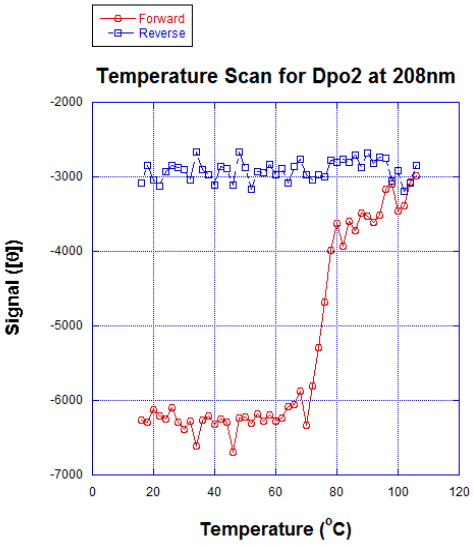

Figure 6. Temperature dependent scan for Dpo2 in the forward and reverse direction at a constant wavelength of 208 nm.

The trends for both the wavelength scans and temperature scan for Dpo3 show a similar pattern. Signal remains relatively constant until after 46 °C, followed by a sudden, sharp increase. After 76 °C, the signal becomes stable once again. Thus, the spectra suggest that that the T_m falls between 46 °C and 76 °C. Analysis of the data obtained by CD spectroscopy suggests that the T_m for Dpo3 is approximately 59 °C, given by the midpoint of the transition.

The spectra obtained from wavelength scans for Dpo2 showed a stable signal from 16 °C to 66 °C, followed by a sudden increase in molar ellipticity at 76 °C after which the signal remained relatively stable (Figure 5). The temperature dependent spectrum for Dpo2 shows a stable signal from 16 °C to about 70 °C, followed by a sharp increase in molar ellipticity from 70 °C to 80 °C. From 80 °C to 106 °C, there is a slow and steady increase in molar ellipticity.

The trends in these spectra are in agreement. The wavelength scans reveal that a significant change in molar ellipticity occurs between 66 °C and 76 °C. The temperature

spectrum agrees, showing a sharp increase that begins at roughly 70 °C and continues to 80 °C before becoming relatively stable. There is a slight difference in signal between the 76 °C and 86 °C species on the wavelength spectra, but there is no discernible difference between the 86 °C, 96 °C, and 106 °C species. These results suggest that the T_m of Dpo2 falls between 66 °C and 86 °C. Analysis of the data obtained from CD spectroscopy reveals that the T_m of Dpo2 is approximately 76 °C, given by the midpoint of the transition.

Discussion

Based on the data reported by our lab for Dpo4, it is noticeable that the T_m value determined by CD spectroscopy for Dpo3 is unusually low. It is reported that the physiological temperature of *Sulfolobus solfataricus* is approximately 80 °C, which is significantly higher than the T_m values reported for Dpo2 and Dpo3 by these studies [7]. Furthermore, a paper published by the Trakselis group in early 2012 dealing with thermostability studies with Dpo3 from *S. solfataricus* estimated the T_m for Dpo3 at about 94 °C using CD spectroscopy (Figure 7) [8]. Both of these accounts conflict with the T_m value we determined for Dpo3 via CD spectroscopy.

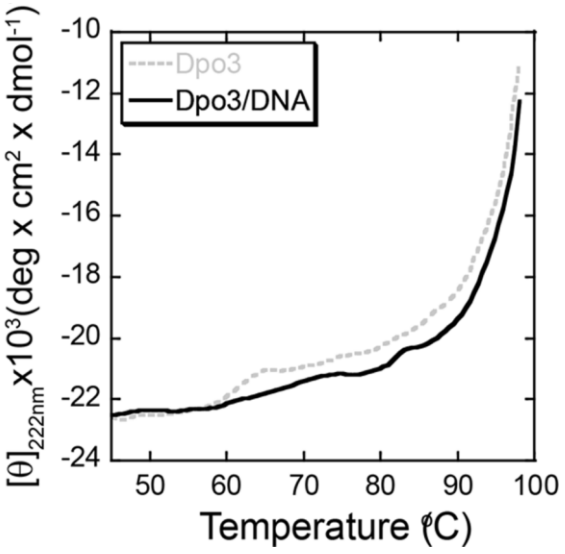

Figure 7. Temperature dependent scan for Dpo3 reported by the Trakselis group

However, plasmids for the expression and purification of Dpo3 were obtained from the Guengerich group [2]. Thermostability studies conducted on Dpo3 by the Guengerich group yielded results similar to our own CD spectroscopy studies. The Guengerich lab conducted temperature-dependent single nucleotide incorporation assays involving the addition of the correct dNTP to an undamaged DNA substrate. Results showed that the optimal reaction

temperature for Dpo3 was 50 °C, well below the physiological temperatures of *Sulfolobus solfataricus* [2]. Furthermore, the Guengerich group reports that the activity of Dpo3 decreased suddenly at temperatures higher than 60 °. Beyond 70 °C, Dpo3 had no activity [2]. This data is supported by our own CD spectroscopy data for Dpo3.

Recently, we have received Dpo3 plasmids from the Trakselis group. I intend to perform additional CD spectroscopy studies to confirm the results obtained by the Trakselis group. If Trakselis' result is reproducible, then there is another issue to be resolved. Which result is correct? Is the data reported by the Guengerich group closer to the real picture, or is the data from the Trakselis group more correct? I hypothesize that the result obtained by the Trakselis group, if reproducible, is more accurate, due to the reported physiological temperature of *S. solfataricus*. It is not evolutionarily logical for *S. solfataricus* to produce a protein that is dead at physiological temperatures.

As with our CD spectroscopy analysis of Dpo3, the T_m value determined for Dpo2 is slightly lower than expected, although the difference is much smaller than with Dpo3. Again, considering a reported physiological temperature for *S. solfataricus* of approximately 80 °C, it seems unlikely that Dpo2 would begin unfolding at temperatures just below 80 °C [6].

The plasmid for Dpo2, like Dpo3, was received from the Guengerich group. The protein expressed by the Dpo2 plasmid was reported to be the carboxy terminus of Dpo2, lacking 65 amino acids on the amino terminus [2]. As with Dpo3, the Guengerich group reported that activity of Dpo2 dropped at 60 °C and was destroyed completely at 70 °C [2]. CD spectroscopy studies conducted in our lab reveal that the structure of Dpo2 is stable up to at least 70 °C.

It is possible the result obtained for Dpo2, unlike that for Dpo3, is reasonably accurate. The Dpo2 expressed by the plasmid received from the Guengerich lab is the 561 amino acid carboxy terminus of Dpo2. Perhaps the 65 amino acid amino terminus that is missing from the

full-length protein is involved in stabilizing structural interactions in full-length Dpo2. In the absence of the 65 amino acid residues on the N-terminus, Dpo2 may be less thermostabile. To answer this question, CD spectroscopy studies for the amino terminus of Dpo2 as well as full-length Dpo2 may be employed in the future.

II. Purification of Human DNA Polymerase ε

Background

Literature published recently by the Pursell group suggested that mutation of residues 275-277 (DIE) to AIA will result in an exonuclease-deficient Pol ε [9]. Using mutagenesis and a Qiagen plasmid purification kit, we were successful in obtaining a plasmid coding for the D275A/E277A double mutant. The plasmid was then transformed in Rosetta cell lines for protein expression. A single colony was collected and stored in 65% glycerol at -80 °C. The following section will discuss an approach that was applied to the purification of the Pol ε double mutant.

Temperature and Induction Time Optimization for Purification of Exonuclease-Deficient Pol ε

A small sample was taken from a stock (A) of protein-containing cells in glycerol and added to 125 mL LB containing 50 μg/mL ampicillin. The culture was allowed to grow for overnight at 37 °C and was then transferred to a flask containing 1.2 L LB and 25 μg/mL ampicillin. The culture was allowed to continue growing at 37 °C until it reached an O.D. of 0.6. The culture was then divided evenly between two 2 L flasks, resulting in two identical 600 mL cultures. One culture was cooled to 25 °C on ice and then induced by 0.5 mM IPTG. The other culture was cooled to 18 °C on ice and also induced by 0.5 mM IPTG. Both cultures were allowed to continue growing at their respective temperatures of 25 °C and 18 °C with 150 mL of each culture being harvested at 2 hours, 4 hours and 6 hours. Additionally, 1 mL post-induction samples of each culture were collected at each time point and diluted to an O.D. 600 of 1.0. These samples were collected to determine if some of the desired protein product was insoluble. The remaining 150 mL of the two cultures were allowed to grow at their respective temperatures

overnight. The harvested pellets were lysed via French press and cleared by ultracentrifugation at 35000 rpm for 40 minutes. The above process was performed a second time simultaneously using a different glycerol stock (B) of exonuclease-deficient Pol ε.

Following the experiment, the samples were separated on two SDS PAGE gels. The first gel comprises of cleared lysate samples obtained via lysis and ultracentrifugation of harvested 150 mL samples (Figure 8A). The second gel comprises of the 1 mL samples that were collected, pelleted, and resuspended at each time point (Figure 8B). Results from SDS PAGE show that stock B better-expressed Pol ε relative to stock A, and that expression was highest when carried out at 18 °C for 4 to 6 hours. production of soluble protein.

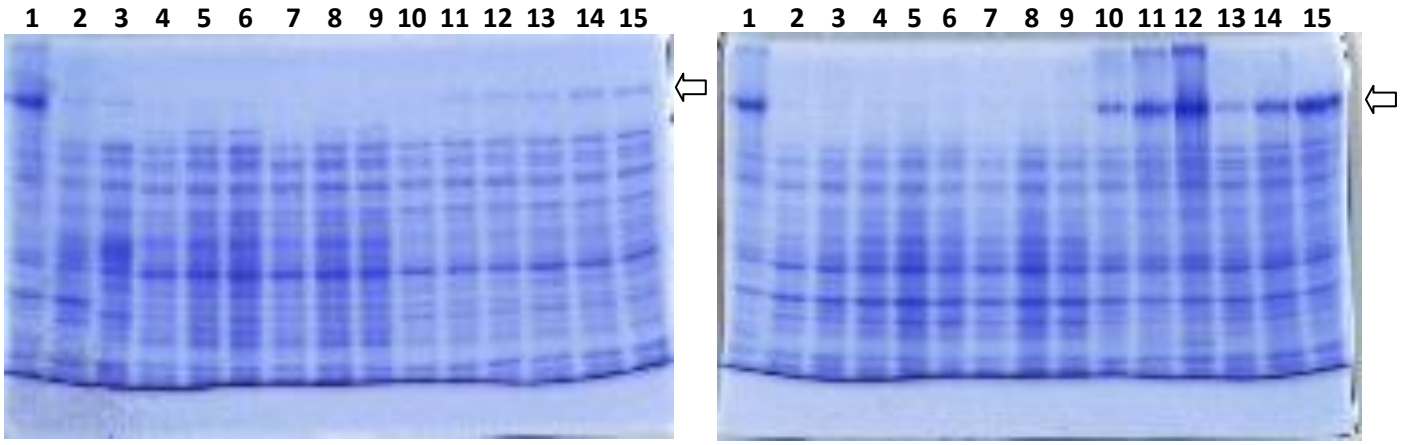

Figure 8A (left). Soluble samples that were harvested, lysed, and cleared by ultracentrifugation. Lane 1 is a DNA size marker; lanes 2 and 3 leaky induction samples collected before and after harvesting, respectively; lanes 4-6 are cells from stock A grown at 25 °C and harvested at 2, 4, and 6 hours after induction; lanes 7-9 are cells from stock A grown at 18 °C and harvested at 2, 4, and 6 hours after induction; lanes 10-15 are stock B samples that were grown at temperatures and times identical to the stock A cells in lanes 4-9. Epsilon is marked by an arrow
Figure 8B (right). Insoluble samples that were pelleted and resuspended prior to loading. Lane 1 is a DNA size marker; lanes 2 and 3 are pre- and post-induction samples; lanes 4-15 were obtained from conditions identical to those in lanes 4-15 and were in figure B (left), and were pelleted and resuspended prior to loading. Epsilon is marked by an arrow.

Expression and Purification of Exonuclease-Deficient Pol ε

A small sample was taken from a stock of protein-expressing cells in glycerol and added to 125 mL of LB media containing ampicillin at 50 μg/mL. The culture was allowed to grow at 37 °C overnight. The following day, 1.2 L of LB media and 25 μg/mL ampicillin were preheated to 35 °C. The overnight culture was then divided evenly to inoculate each flask of 1.2 L media. The cultures were incubated at 37 °C for 3 hours until the O.D. 600 had reached 0.6. The cultures were then chilled to 18 °C on ice. The cultures were induced by the addition of 0.25 mM IPTG. The O.D. 600 at the time of induction was 0.67. Incubation was allowed to continue for 4 hours at 18 °C. At an O.D. 600 of 1.0, the cells were harvested by centrifugation at 3000 rpm for 10 minutes. The cell pellets were frozen in liquid nitrogen and stored overnight at 80 °C. The following day, the cell pellets were thawed on ice and resuspended in a 50:50 mixture of lysis buffer containing 300 mM Tris HCl pH 7.8 at 4°C, 100 mM NaCl, 20 mM K_2SO_4, 0.5 mM EDTA, 2 mM PMSF, and protease inhibitor cocktail. The cells were lysed via French pressure cell press and then clarified by ultracentrifugation at 35,000 rpm for 40 minutes. The cleared lysate was incubated in the presence of NiMAC beads and 50% lysis buffer for 2 hours. The NiMAC beads were packed into a column and washed with 50% lysis buffer for 25 column volumes (CV). The NiMAC beads were then washed more stringently for 10 CV in 50% lysis buffer containing 5 mM imidazole. The protein was eluted from the NiMAC beads by 50% lysis buffer and a linear gradient from 5 mM to 500 mM imidazole. The protein-containing fractions were pooled and dialyzed overnight against 2 L dialysis buffer containing 150 mM Tris HCl pH 7.8 at 4 °C, 200 mM NaCl, 20 mM K_2SO_4, and 2 mM DTT. The dialyzed protein was concentrated via centrifugation using a centriprep with 10 kDa cutoff. The concentrated protein was loaded onto an S200 size exclusion column and eluted in dialysis buffer. The protein-containing fractions were pooled and diluted to a salt concentration of 100 mM and loaded onto a

Mono S column for purification by ion exchange. The protein was eluted from the column at a salt concentration of 220 mM. The purity of the protein was confirmed via SDS-PAGE. Finally, the pure protein-containing fractions were pooled and concentrated using a 10 kDa centriprep. 100 μg of pure protein were obtained.

III. Characterization of Human DNA Polymerase ε

Background

Following successful purification of the exonuclease-deficient Pol ε, we performed experiments to be sure that the mutation did not inhibit polymerase activity. Additionally, we performed a single nucleotide misincorporation assay to obtain an estimate for misincorporation rate. With this estimate, we will be able to design experiments for the determination of k_d for nucleotide misincorporation in the near future.

Recently we received from our collaborator a 45mer DNA template and several 20mer DNA primers with increasing NTP content. The primers are labeled R_n, where n represents the number of NTPs at the 3' end of the 20mer primer. All of the primers are 20 units long; therefore the primer labeled R_2 contains 18 dNTPs and ends in 2 NTPs. All of the primers are capable of annealing correctly and completely with the provided 45mer DNA template. The 45mer template contains no NTPs. Our lab was charged with the task of performing exonuclease assays using wild-type human DNA polymerase ε and the provided substrates. The following section will outline the procedures followed in preparing and conducting the experiments in addition to the data obtained as of current.

Materials and Methods

Radiolabeling and annealing: A 21mer DNA primer was 5' ^{32}P radiolabeled by using T4 kinase. The radiolabeled primer was then purified from free gamma-ATP using Bio-Spin™ Tris Columns. The primer was then annealed to a 41mer D1 DNA template in a 1:1.5 ratio by heating at 95 °C for five minutes and cooling to room temperature overnight. 2 μM of annealed

Table 1: DNA Substrates*	
D1 21/41mer	5’-CGCAGCCGTCCAACCAACTCA 3’-GCGTCGGCAGGTTGGTTGAGTAGCAGCTAGGTTACGGCAGG-5’
R_0	5’-CCTCTTCGCTATTACGCCAG-3’
R_1	5’-CCTCTTCGCTATTACGCCA**G**-3’
R_2	5’-CCTCTTCGCTATTACGCC**AG**-3’
R_3	5’-CCTCTTCGCTATTACGC**CAG**-3’
R_4	5’-CCTCTTCGCTATTACG**CCAG**-3’
R_5	5’-CCTCTTCGCTATTAC**GCCAG**-3’
R_n 45mer	3’-GGAGAAGCGATAATGCGGTCGACCGCTTTCCCCCTACACGACGTT-5’
*Nucleotides in **bold** are ribonucleotides	

primer/template substrate was obtained. The R_n primers for the exonuclease assays were prepared in an identical fashion. Radiolabeled primers ranging from 1.5-2 μM were obtained from each reaction. The DNA/RNA primers were then annealed to the 45mer DNA template in 1:1.5 primer:template ratio at 75 °C (Table 1).

Activity assay: General activity assays were performed under both pre-steady state and steady state conditions. Two solutions were prepared for each reaction. For the pre-steady state reaction, Solution A contained 50 mM Tris HCl pH 7.4 at 37 °C, 1 mM DTT, 100 μM BSA, 10 μM EDTA, 240 nM annealed 21/41mer D1, and 80 nM exonuclease-deficient Pol ε. Solution B was identical to Solution A, but did not contain DNA or enzyme, and included 200 μM dTTP and 16 mM $MgCl_2$. Solutions A and B for the steady state reaction were prepared in an identical manner, but contained 480 nM 21/41mer D1 and only 6 nM exonuclease-deficient Pol ε. Reactions were started by mixing equal volumes of A and B and were quenched at various time points by the addition of 0.37 M EDTA. The assays were performed at 37 °C. The pre-steady state assay was performed on rapid quench, while the steady state assay was performed manually. An additional manual quench was performed using identical conditions, but with the addition of dTTP in Solution B omitted in order to test for residual exonuclease activity.

Products were separated via PAGE and visualized using a phosphorimager. Data was quantitated using computer software.

Misincorporation assay: Two separate solutions were prepared. Solution A contained 50 mM Tris HCl pH 7.4 at 37 °C, 1 mM DTT, 100 μM BSA, 10 μM EDTA, 30 nM annealed 21/41mer D1, and 300 nM exonuclease-deficient Pol ε. Solution B was identical, but omitted DNA and enzyme, and included 16 mM $MgCl_2$ and 200 μM dATP. Reactions were performed using a rapid quench apparatus. The reactions were started by the mixing equal volumes of solutions A and B. At various time points, the reactions were quenched by the addition of 0.37 M EDTA. The assay was performed at 37 °C. Products were separated via PAGE. Gel images were obtained using a phosphorimager and quantitated using computer software.

Exonuclease assays with DNA/RNA hybrid substrates: Two separate solutions were prepared for each reaction. Solution A contained 50 mM Tris HCl pH 7.4 at 37 °C , 1 mM DTT, 100 μM BSA, 10 μM EDTA, 50 mM NaCl, 30 nM annealed 20/45mer, and 300 nM wild-type Pol ε. The B solution was identical, but contained no DNA or Pol ε and included 16 mM $MgCl_2$. Reactions were performed using a rapid quench apparatus. The reactions were started by the mixing equal volumes of solutions A and B. At various time points, the reactions were quenched by the addition of 0.37 M EDTA. All assays were performed at 37 °C. Products were separated via PAGE. Gel images were obtained using a phosphorimager and quantitated using computer software. Data from each experiment was fit to a single exponential decay equation, both with and without an intercept, and a double exponential decay equation.

Results

Activity assay: The pre-steady state incorporation assay testing the activity for exonuclease-deficient Pol ε showed an initial burst rate of 72 s^{-1} (Figure 10). The accompanying

steady state reaction showed an incorporation rate of 0.01 s^{-1} (Figure 11). Using the rate of 72 s^{-1} as an estimate, we will perform a full k_d analysis of correct nucleotide incorporation.

Furthermore, we performed a brief exonuclease assay using the Pol ε exonuclease-deficient mutant. By 40 minutes after the reaction was started, exonuclease activity was detected (data not shown). For our purposes, it is likely that the D275A/E277A double mutant of Pol ε is suitable for kinetic studies. However, we are interested in performing a triple mutation in hopes of totally eradicating exonuclease activity. The details for designing the mutant will be discussed later in the thesis.

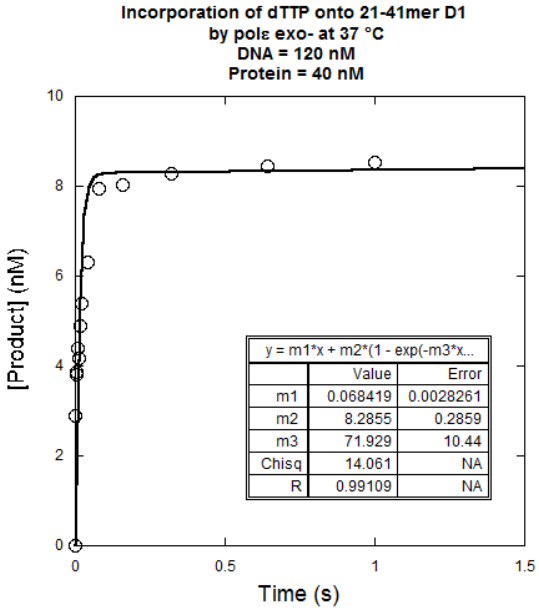

Figure 10. Pre-steady state kinetic data for correct nucleotide incorporation by Pol ε exo-minus.

Misincorporation assay: Data collected from the misincorporation assay involving the insertion of the incorrect nucleotide (dATP) across from a template base at the 22nd position on the template described an incorporation rate of 1 s^{-1} (Figure 12). This figure is in agreement with a previous study in which misincorporation of dTTP across from T showed a reaction rate of 1.1 s^{-1} [9]. Compared with the reaction rate of correct nucleotide incorporation, this data suggests that Pol ε is more likely to readily incorporate

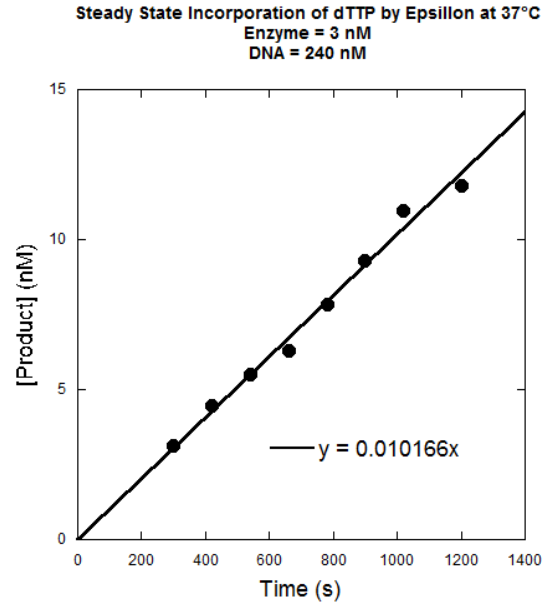

Figure 11. Steady state kinetic data for correct nucleotide incorporation by Pol ε exo-minus.

the correct base rather than an incorrect base.

Exonuclease assays with DNA/RNA hybrid substrates: Currently, degradation assays for only R_2, R_3, and R_5 substrates have been performed. Neither of the equations used to fit the data to a curve were able to elucidate a clear trend relating the three R_n substrates tested (Tables 2-4 and Figures 13-16). However, all three fits illustrate a drop in degradation rate from R_2 to R_3 and a subsequent increase in rate from R_3 to R_5. We hypothesize one of two scenarios: First, it is possible that, once there are so many NTPs at the 3' end of the primer, a further addition of NTPs has no additive effect on reaction rate. Secondly, it is possible that a large number of NTPs at the end of a primer may actually inhibit exonuclease-mediated degradation. That is, a small number of NTPs at the end of a primer may initially facilitate degradation, but a larger number may inhibit exonuclease activity. To test these hypotheses, the remaining R_n substrates will be studied using identical experimental and data analysis procedures.

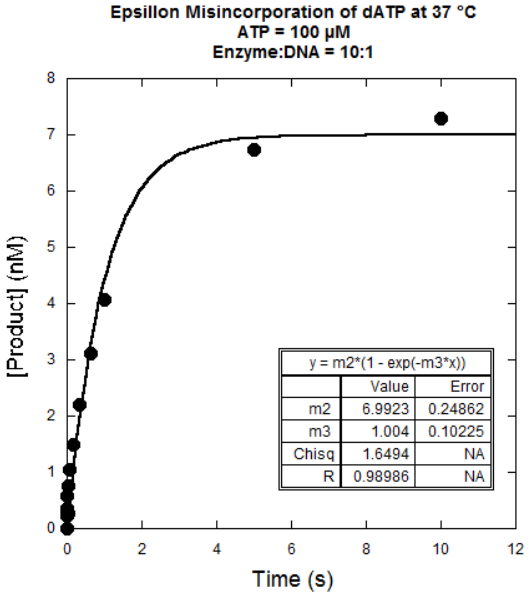

Figure 12. Data for misincorporation of dATP across from A on the template.

Table 2: Comparison of Degradation Rates of R_n Substrates Fit to Single Exponential Decay			
	R_2	R_3	R_5
Rate (s^{-1})	0.18	0.08	0.2
Amplitude (%)	69.3	94.7	91.3
Table 3: Comparison of Degradation Rates of R_n Substrates Fit to Single Exp. Decay w/ Intercept			
	R_2	R_3	R_5
Rate (s^{-1})	0.63	0.15	0.29
Amplitude (%)	56.7	69.3	79
Table 4: Comparison of Degradation Rates of R_n Substrates Fit to Double Exponential Decay			
	R_2	R_3	R_5
Slow Rate (s^{-1})	0.12	0.034	0.005
Slow Amplitude (%)	56	54	15.3
Fast Rate (s^{-1})	29.3	0.23	0.3
Fast Amplitude (%)	45.3	42.7	77.3

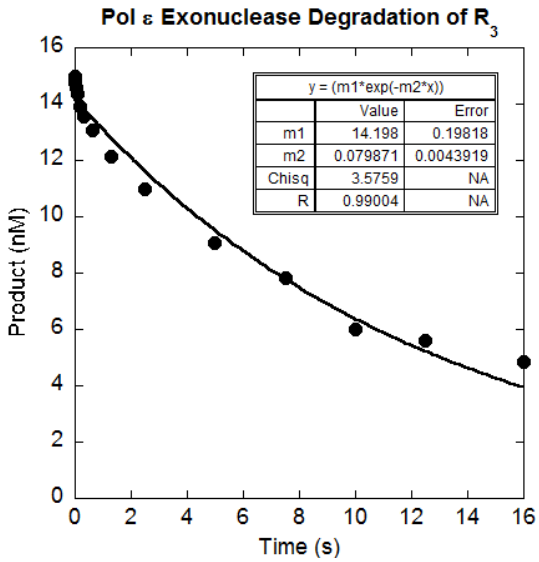

Figure 13. Pol ε exo-minus degradation of R_3 substrate fit to a single exponential decay equation.

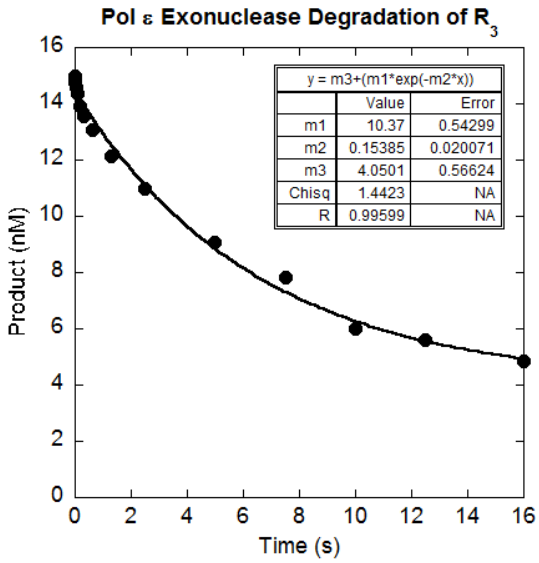

Figure 14. Pol ε exo-minus degradation of R_3 substrate fit to a single exponential decay equation with an intercept.

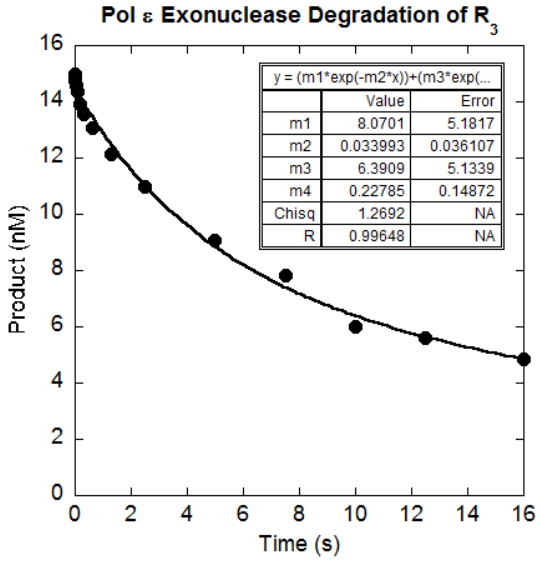

Figure 15. Pol ε exo-minus degradation of R_3 substrate fit to a double exponential decay equation.

IV. Knocking Out Exonuclease Activity in Dpo2 and Pol ε

Background

For full characterization of *S. solfataricus* Dpo2 and Dpo3 and Human DNA Pol ε, it is important to create exonuclease-deficient mutants. All three of these DNA polymerases have 3'-5' exonuclease proofreading activity. The rates of polymerase activity may be difficult to accurately determine if exonuclease activity is present and causing even small amounts of product degradation. Exonuclease activity will also interfere with the determination of nucleotide misincorporation rates. That is, exonuclease activity rates are often higher than the incorrect base incorporation rates in order to prevent the formation of mispaired DNA. By eliminating exonuclease activity, we can see the true rate of incorrect base incorporation and accurately assess the mutagenicity of both the error-prone polymerases from *S. solfataricus* and the highly accurate human Pol ε. Furthermore, our lab is interested in studying how polymerase enzymes interact with damaged DNA substrates. Like with misincorporation, incorporation across from damaged sites is often slower than 3'-5' exonuclease activity, thus preventing the replication of damaged DNA.

Family B DNA polymerases typically contain three highly conserved exonuclease motifs [10]. Exo-motif I is characterized by DxE residues, where x represents any amino acid. Exo-motif II is marked by NxxxF/YD residues, and exo-motif III is distinguished by YxxxD residues [10]. These motifs usually contain residues that are essential for coordinating metal ions that catalyze exonuclease activity. As a strategy for knocking out exonuclease activity, it is useful to compare the primary structure of the polymerase of interest to all other known enzymes and locate a polymerase with a similar sequence that has been previously mutated to remove exonuclease activity.

Exonuclease-Deficient Pol ε

A paper published in 2011 by the Pursell group demonstrated that exonuclease activity for human DNA polymerase ε could be knocked out completely by the mutation of resides 275-277 (DIE) to AIA [9]. We performed mutagenesis on wild-type Pol ε to create the D275A/E277A double mutant. Activity assays demonstrated that some exonuclease activity was still present and was causing product degradation by 40 minutes after Pol ε and the 21/41 DNA substrate were mixed. This result is contrary to the claim by the Pursell group that the aforementioned double mutation was sufficient for completely eliminating exonuclease activity (Figure 16) [9].

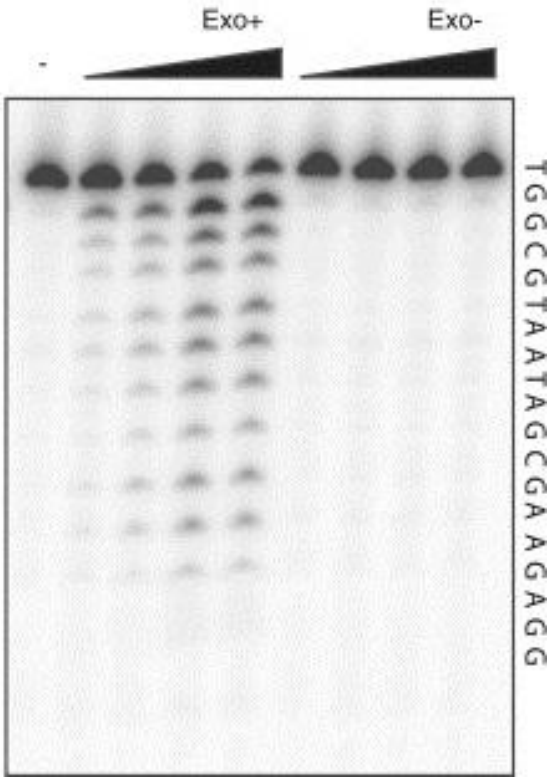

Figure 16. Gel published by Pursell group showing efficiency of the D275A/E277A double mutant in knocking out exonuclease activity. Lane 1contains no enzyme. Lanes 2-5 contain wild-type Pol ε in increasing concentrations. Lanes 6-9 contain exo-deficient Pol ε in increasing concentrations.

We discovered that the publication by the Pursell group failed to mention how long the exonuclease reactions were allowed to continue prior to stoppage by the addition of 95% formamide [9]. Given both the ambiguity of the Pursell group's experiment and the results from our own exonuclease activity assay, we have decided to select a third residue for a triple mutant in hopes of completely removing exonuclease activity altogether.

Alignments between human Pol ε, yeast Pol ε, *S. solfataricus* Dpo1, and yeast Pol δ revealed a highly-conserved DIE motif in all four enzymes (Figure 17). Downstream from these three residues is another highly conserved aspartate residue. In Dpo1, the triple mutant D231A/E233A/D318A was successful in removing exonuclease activity [11]. We believe that the high conservation of these residues across these four essential polymerases is a good

indication that the third residue in human Pol ε (D369) is also involved in coordinating metal ions involved in exonuclease activity. Thus, we propose that a triple mutant of Pol ε, D275A/E277A/D369A will be successful in totally destroying exonuclease activity.

```
yEpsilon        DKDIRVGKWYKVTQQG---FIEDTRK---IAFADPVVMAFDIETTKPPL-KFPDSAVDQI 308
hEpsilon        DLKIHVAHWYNVRYRGNAFPVEITRRDDLVERPDPVVLAFDIETTKLPL-KFPDAETDQI 294
Dpo1            KDVEEIKKAFADSDEMTRQMAVDWLPIFETEIPKIKRVAIDIEVYTPVKGRIPDSQKAEF 250
yDelta          EPNNRVSSCQLEVSINYRNLIAHPAEGDWSHTAPLRIMSFDIECAGRIG-VFPEPEYDPV 339
                .   .:                            .    :::***        :*:.

yEpsilon        VRPTVISTFNGDFFDWPFIHNRSKIHGLDMFDEIGFAPDAEGEYKSSYCSHMDCFRWVKR 428
hEpsilon        TKPTIMVTYNGDFFDWPFVEARAAVHGLSMQQEIGFQKDSQGEYKAPQCIHMDCLRWVKR 414
Dpo1            YP--IVLTFNGDDFDLPYIYFRALKLGY-FPEEIPIDVAGKDEAKYLAGLHIDLYKFFFN 362
yDelta          VDPDVIIGYNTTNFDIPYLLNRAKALKVNDFPYFGRLKTVKQEIKESVFSSKAYGTRETK 452
                    ::  :*   ** *::  *:          :      : * *              .
```

Figure 17. Alignment between yeast Pol ε, human Pol ε, *S. solfataricus* Dpo1, and yeast Pol δ. Highlighted in yellow are residues that are highly conserved across the four enzymes. Highlighted in green are mutations in Dpo1 that created exonuclease-deficiency. Highlight in blue are the residues we intend to mutate to completely abolish exonuclease activity in Pol ε.

Exonuclease-Deficient Dpo2

To determine which residues should be mutated to create exonuclease-deficient Dpo2, we performed a BLAST search to locate enzymes with similar sequences which have also been mutated for exonuclease deficiency. BLAST searches revealed that Dpo2 is a highly unique enzyme and does not share a high degree of sequence similarity with most known enzymes. We were unable to locate a similar polymerase for which exonuclease activity has been removed. However, we were able to locate DNA Polymerase B2 from *Sulfurisphaera ohwakuensis*, a protein with reasonable sequence similarity to Dpo2 (68% identity) for which an exonuclease domain motif was defined. In a paper examining three B family DNA polymerases from *S. ohwakuensis*, the Horiuchi group mapped out the exo-motif III (YxxxD) for Pol B2. They report that the sequence of Pol B2 did not contain either exo-motif I or exo-motif II [10]. They also compare Pol B2 with *S. solfataricus* Dpo2, stating that both have similar amino acid substitutions in the polymerase motifs [10]. As such, Pol B2 may serve as a helpful model in our studies of

Dpo2. The alignment between *S. ohwakuensis Pol* B2 and *S. solfataricus* Dpo2 shows that the exo-motif III (YxxxD) of Dpo2 contains a nucleotide substitution from D to E at position 230 (Figure 18). However, this glutamate substitution conserves the acidic nature of the aspartate residue in the exo-motif. Position 227 in the Dpo2 sequence also contains a glutamate residue. It is possible that these two residues are responsible for coordinating metal ions involved in Dpo2 exonuclease activity. However, slightly downstream from this motif from positions 234 to 237 is a DIAE sequence, which is similar to the DxE exo-motif I commonly found in B family DNA polymerases. We suspect that the D234 and E237 residues may be involved in metal ion coordination. As such, we intended to create a D234A/E237A double mutant in an attempt to knockout exonuclease activity.

```
gi|7416105|dbj|BAA93702.1|          FLRWYDGCTD--CYEINGE----RVYDLDDFEADVVECYGFPCNKIKAQV 163
gi|14424453|sp|Q07635.2|DPOL2_      PLDWYGESLKGKVFEVKINNEVRRFYEKPEVEVDIAECLGEACNYVKSNV 250
                                     * **. . .   :*:: :    *.*:  :.*.*:.** * .** :*::*
```

Figure 18. Alignment between *S. solfataricus* Dpo2 and *Sulfurisphaera ohwakuensis* DNA Polymerase B2. Highlighted in green is the conserved exo-motif III. Highlighted in yellow are the acidic gultamate residues that align with the aspartate residues in *S. ohwakuensis*. Highlighted in blue are the residues that closely resemble the DIE exonuclease motif. These are the residues that are intended for mutation to AIAA.

Conclusion

In Chapter 1, we used circular dichroism spectroscopy to characterize the thermostability of DNA Polymerase II and III from *Sulfolobus solfataricus*, a thermophilic archaea. We obtained T_m values for Dpo2 and Dpo3 that were both lower than the T_m values reported by our lab for Dpo4 and in contradiction to the data reported by the Trakselis group in a similar study on Dpo3. However, the research group from which we received the plasmids for expressing Dpo2 and Dpo3 reported thermostability figures obtained from kinetic analysis that were similar to our own observed values. It is evident that there lies a discrepancy between the plasmids we expressed and those expressed by the Trakselis lab. We hypothesize that the results reported by the Trakselis group are accurate and that the plasmids we received may not actually express full-length Dpo3, but rather some significantly more-thermolabile structure. We hypothesize that the low thermostability of Dpo2 can be attributed to the 65 amino acid truncation on the protein's amino terminus.

In Chapter 2, we optimized the temperature and induction time for expressing exonuclease-deficient Pol ε. We determined that growth at 18 °C for 4-6 hours results in the best expression. Furthermore, we determined during the same experiment that much of the protein remains insoluble following induction with IPTG. Finally, we analyzed the expression and purification of exonuclease-deficient Pol ε.

In Chapter 3, we performed a number of assays using purified exonuclease-deficient Pol ε. The rate of correct incorporation of a single nucleotide was significantly higher than single misincorporation. Additionally, exonuclease activity was still present after 40 minutes, indicating that exonuclease activity was not completely destroyed by the D275A/E277A double mutation. Furthermore, we performed exonuclease assays with wild-type Pol ε and DNA/RNA hybrid primers. There is no obvious pattern for the three substrates we tested, but we hypothesize that

the exonuclease activity of Pol ε may be inhibited by a large number of ribonucleotides at the 3' end of the primer.

In Chapter 4, we performed sequence alignments for Pol ε and Dpo2. For Pol ε, we determined that the double mutant still contained residual exonuclease activity, so a third residue would have to be mutated in order to completely remove activity. The alignment was performed between Pol ε, yeast Pol ε, *S. solfataricus* Dpo1, and yeast Pol δ. From this alignment, we determined that D369 would be the most logical residue to modify as this residue aligns with a residue in Dpo1 that was mutated to successfully eliminate exonuclease activity. Therefore, we will be testing a D275A/E277A/D369A triple mutant for Pol ε. For Dpo2, we were unable to find a large range of highly similar proteins. However, DNA polymerase B2 from *Sulfurisphaera ohwakuensis* did have reasonably similar sequence identity and one of its exonuclease domains was identified. Using this information, we will be testing a D234A/E237A double mutant for Dpo2.

This thesis is a summary of the work I have done since winter 2012 in the Suo lab. I was aided by Jason Fowler, Dave Taggart, and Vineet Gaur in my studies, and they have been a great influence in helping me to develop as an effective scientist. I would also like to thank Dr. Zucai Suo for his support and encouragement. I am looking forward to continuing my graduate studies in the Suo lab in autumn 2012.

References

1. Pohjola, S.K., et al., *DNA binding of polycyclic aromatic hydrocarbons in a human bronchial epithelial cell line treated with diesel and gasoline particulate extracts and benzo[a]pyrene.* Mutagenesis, 2003. **18**(5): p. 429-38.
2. Choi, J.Y., et al., *Roles of the four DNA polymerases of the crenarchaeon Sulfolobus solfataricus and accessory proteins in DNA replication.* J Biol Chem. **286**(36): p. 31180-93.
3. Kunkel, T.A., *DNA replication fidelity.* J Biol Chem, 2004. **279**(17): p. 16895-8.
4. Baase, W.A., et al., *DNA models of trinucleotide frameshift deletions: the formation of loops and bulges at the primer-template junction.* Nucleic Acids Res, 2009. **37**(5): p. 1682-9.
5. Chui, G. and S. Linn, *Further characterization of HeLa DNA polymerase epsilon.* J Biol Chem, 1995. **270**(14): p. 7799-808.
6. Unpublished source, o.f.i.c.
7. Zaparty, M., et al., *"Hot standards" for the thermoacidophilic archaeon Sulfolobus solfataricus.* Extremophiles. **14**(1): p. 119-42.
8. Bauer, R.J., M.T. Begley, and M.A. Trakselis, *Kinetics and fidelity of polymerization by DNA polymerase III from Sulfolobus solfataricus.* Biochemistry. **51**(9): p. 1996-2007.
9. Korona, D.A., K.G. Lecompte, and Z.F. Pursell, *The high fidelity and unique error signature of human DNA polymerase epsilon.* Nucleic Acids Res. **39**(5): p. 1763-73.
10. Iwai, T., et al., *Sequence analysis of three family B DNA polymerases from the thermoacidophilic crenarchaeon Sulfurisphaera ohwakuensis.* DNA Res, 2000. **7**(4): p. 243-51.

11. Zhang, L., et al., *Polymerization fidelity of a replicative DNA polymerase from the hyperthermophilic archaeon Sulfolobus solfataricus P2.* Biochemistry, 2009. **48**(31): p. 7492-501.